

*Systematic Literature Review***A Systematic Review of the Testing Effect in Learning**

Raquel Eloísa Eisenkraemer

Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre-RS, Brazil

Antônio Jaeger

*Universidade de São Paulo, São Paulo-SP, Brazil*Lilian Milnitsky Stein¹*Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre-RS, Brazil*

Abstract: The retrieval of a given piece of information from memory increases the long-term retention of that information, a phenomenon often called “testing effect”. The current study aimed to select and review articles on the testing effect to verify the extent and importance of this phenomenon, bringing the main results of recent research. To accomplish this, a systematic review of articles on this subject published between 2006 and 2012 was conducted, a period in which there was an acute increase in the amount of publications on this subject. The articles were searched in the databases Web of Science, PubMed and PsycINFO. The results, which were organized according to test format (recall and recognition tests), demonstrated that tests can be remarkably beneficial to the retention of long-term memories. A theoretical explanation regarding the cognitive processes involved in this phenomenon still needs to be developed and tested. Such explanation would have important implications for the development of efficient educational practices.

Keywords: memory, information retrieval, learning

Uma Revisão Sistemática Sobre o Efeito de Testagem na Aprendizagem

Resumo: Recordar informações previamente memorizadas aumenta a retenção destas informações na memória, um fenômeno chamado “efeito de testagem”. O presente estudo objetivou selecionar e avaliar artigos sobre o efeito de testagem, a fim de verificar a extensão e importância desse fenômeno, trazendo os principais resultados de pesquisas recentes. Para isto, realizou-se uma revisão sistemática de artigos publicados entre 2006 e 2012, período em que houve um aumento expressivo na quantidade de publicações sobre este tema. A busca dos artigos foi desenvolvida nas bases de dados Web of Science, PubMed e PsycINFO. Os resultados, organizados quanto ao formato de teste (recordação e reconhecimento), indicaram que a realização de testes produz efeitos notavelmente benéficos para a retenção de memórias de longo prazo. Entretanto, uma explanação teórica referente aos processos cognitivos envolvidos neste fenômeno ainda precisa ser desenvolvida e testada uma vez que tal explicação possuiria importantes implicações para o desenvolvimento de práticas educacionais eficazes.

Palavras-chave: memória, recuperação da informação, aprendizagem

Una Revisión Sistemática del Efecto de Prueba en el Aprendizaje

Resumen: Recordar información previamente almacenada aumenta la retención posterior de esta información, un fenómeno llamado de “efecto de prueba”. En el presente estudio se buscó seleccionar y evaluar artículos sobre el efecto de prueba para verificar el alcance y la importancia de este fenómeno, presentando los principales resultados de investigaciones recientes. Para eso, fue llevada a cabo una revisión sistemática de artículos publicados entre 2006 y 2012, periodo en el cual se observó un aumento significativo en el número de publicaciones respecto al tema. La búsqueda de los artículos se realizó en las bases de datos Web of Science, PubMed y PsycINFO. Los resultados, organizados por tipo de prueba (recuerdo y reconocimiento), dan evidencia de que las pruebas producen efectos significativos en beneficio de la retención de información a largo plazo. Por otro lado, una explicación teórica referente a los procesos cognitivos involucrados en este fenómeno tiene que ser desarrollados y evaluados, una vez que se tenga tal explicación se tendrán importantes implicaciones para el desarrollo de prácticas educativas eficientes.

Palabras clave: memoria, recuperación de la información, aprendizaje

In educational contexts, teachers generally consider tests or exams as assessment devices to measure the students' learning of certain contents over a specific time period (Roediger & Butler, 2011). In different circumstances, tests are not applied frequently, like once for each set of

contents or at the end of the term or semester for example. In addition, both teachers and students tend to see them as a bother (Roediger & Karpicke, 2006b). Recent studies (Moreira-Aguiar et al., 2008), however, have demonstrated that tests positively affect the long-term memory (long-term information retention), suggesting that individuals tested for one kind of material, with successful recall, will better remember this material in the near future. Researchers have called this phenomenon the “testing effect” (Roediger & Karpicke, 2006b).

¹ Correspondence address:

Lilian Milnitsky Stein, Programa de Pós-Graduação em Psicologia da Pontifícia Universidade Católica do Rio Grande do Sul, Avenida Ipiranga, 6681. Prédio 11/sala 940. CEP 90619-900. Porto Alegre-RS, Brazil. E-mail: lilian@puers.br

Recently, interest in investigating the testing effect has increased (Karpicke, 2012). One of the main reasons for this growing interest is the possibility to apply the findings for the benefit and improvement of educational practices. Tests provide gains that go beyond a mere assessment of the contents learned. Recovering information from one's memory favors the modification of the memory trait itself, consequently increasing the possibility of future success in recovering the information learned. Hence, memory tests, that is, recovery practices that evoke the memory of the contents addressed by the teacher can be used as learning tools in the classroom (Halamish & Bjork, 2011; Roediger, McDermott, & McDaniel, 2011). On the opposite, it seems that Educational researchers and professionals are insufficiently aware of these benefits. One reason for this fact can be the lack of studies to explore the benefits of the testing effect in natural teaching environments.

Experiments related to the testing effect normally involve a range of memory tests. Research participants initially study some material (observation of lists of words or images or reading of texts), and then either recall the material studied through recall or recognition memory tests, or simply restudy the material. After different intervals, a final (recall or recognition) memory test is taken to investigate what was learned (Roediger & Butler, 2011), and to examine how the fact of having previously learned certain items, soon after studying the material, influenced the memory performance. Thus, the testing effect is verified by exposing the participants to the study phase (Phase 1), initial test or restudy (Phase 2), and final test about the material studied (Phase 3).

The testing effect has been investigated in countless experiments, using different kinds of materials for the study phase (Halamish & Bjork, 2011), such as isolated words (Hogan & Kintsch, 1971), associated word pairs (Allen, Mahler, & Estes, 1969; Bouwmeester & Verkoeijen, 2011; Estes, 1960; Pashler, Cepeda, Wixted, & Rohrer, 2005), maps (Carpenter & Pashler, 2007; Rohrer, Taylor, & Sholar, 2010), faces (Weinstein, McDermott, & Szpunar, 2011), figures of birds (Jacoby, Wahlheim, & Coane, 2010) and texts (Callender & McDaniel, 2009).

After the study phase, the participants are submitted to initial memory tests for the evocation of previously studied information. These tests can be characterized as recall or recognition tests. In recall memory tests, individuals look for information directly in their memory (Howe, 1991). This means that the subject mentally reconstructs the information that was memorized (Anderson, 2005), like in free recall tests (recalling everything that was studied at an earlier time) and clue-based recall tests (completing gaps). Recognition tests, on the opposite, involve the comparison between information stored in the memory and stimuli presented during the test (Watkins, 1974), that is, participants distinguish between "new" items (not presented earlier) and "old" items (presented earlier)

(Yonelinas, 2002). True and false tests and multiple-choice tests, for example, involve the recognition process.

In Education, often, students are hardly stimulated towards dedication in their cognitive processes and towards the creation of strategies to promote their learning when tests are applied. Research to understand these processes entails important implications for educational practice. The current stage of research about the testing effect and learning reflects the migration from the laboratory to the classroom environment and already shows concrete applications of this effect in the educational context. Therefore, it is important to disseminate current research results on memory to promote the teaching-learning process.

Thus, the aim in this systematic review was to identify, select and assess studies about the testing effect. In the review, evidence was considered from experimental laboratory studies, and mainly from studies conducted in situations that stand closer to the school environment, using material from more complex studies. The data survey was aimed at verifying the extent and importance of the testing effect phenomenon, and at investigating the circumstances in which studies on the testing effect are developed, focusing on whether these studies support the idea that recovery (tests) promotes the long-term retention and, hence, delays the forgetting of the information studied.

Method

To search for scientific references, articles were selected from the databases Web of Science, PubMed and PsycINFO, using the descriptors "testing effect" and "learning", in March 2012. To complement the search, papers were selected through the citations in the initially identified articles. Because of the dramatic increase in the number of publications about the testing effect in recent years, only papers published as from 2006 were included in the review. This period was considered sufficient to demonstrate the trends in the research theme.

To analyze the results by means of the systematic review, the following inclusion criteria were considered: (1) paper published in a scientific journal between 2006 and 2012; (2) empirical paper; (3) comply with the research paradigm about the testing effect; (4) include the descriptors "testing effect" and "learning"; (5) involving healthy participants, that is, no papers with clinical samples were included; (6) not reporting on drugs experiments, and (7) having used verbal stimuli as study material. Any papers that did not comply with these criteria were discarded.

Results

Initially, 524 articles were found that used the descriptors "testing effect" and "learning" in the three databases that were consulted (177 in Web of Science, 152 in PubMed and 195 in PsycINFO). Eight other relevant articles were identified based on other sources, such as a reference

list of investigated articles. After excluding repeated studies, the number dropped to 326 articles.

The abstracts of the articles were assessed based on the inclusion criteria, reaching a total number of 31 papers (Table 1), from which information was surveyed to review the experiments used to manipulate the testing effect. For this systematic literature review, only studies that involved verbal stimuli in the study phase were considered. All of the selected studies ratified the benefits of the testing effect, that is, the use of tests promoted long-term retention and learning. The most relevant results of the papers were grouped according to the test format: recall tests (free recall and cued recall) and recognition tests.

Table 1

Systematic Review Corpus

1. Studies about testing effect with recall tests
1.1. Free recall tests
Blumen and Stern (2011)
Congleton and Rajaram (2011)
Karpicke and Roediger (2007)
Karpicke and Roediger (2010)
Roediger and Karpicke (2006a)
Szpunar, McDermott and Roediger (2008)
Zaromb and Roediger (2010)
1.2. Cued recall tests
Agarwal, Karpicke, Kang, Roediger and McDermott (2008)
Butler (2010)
Butler and Roediger (2007)
Carpenter (2011)
Carpenter, Pashler and Vul (2006)
Carpenter, Pashler, Wixted and Vul (2008)
Chan, McDermott and Roediger (2006)
Hays, Kornell and Bjork (2010)
Karpicke and Bauernschmidt (2011)
Kornell, Hays and Bjork (2009)
McDaniel, Roediger and McDermott (2007)
Toppino and Cohen (2009)
Vojdanoska, Cranney and Newell (2010)
2. Studies about testing effect with recognition tests
Butler, Karpicke and Roediger (2007)
Butler, Karpicke and Roediger (2008)
Butler, Marsh, Goode and Roediger (2006)
Butler and Roediger (2008)
Carpenter, Pashler and Cepeda (2009)
Fazio, Agarwal, Marsh and Roediger (2010)
Marsh, Agarwal and Roediger (2009)
Marsh, Roediger, Bjork and Bjork (2007)
McDaniel, Agarwal, Huelser, McDermott and Roediger (2011)
McDaniel, Howard and Einstein (2012)
Roediger, Agarwal, McDaniel and McDermott (2011)

Studies About the Testing Effect Using Recall Tests

Free recall tests. Among the studies selected for the systematic review, seven papers described experiments using free recall tests. This test basically consists in freely recalling the material presented in the study phase (Blumen & Stern, 2011; Congleton & Rajaram, 2011; Wixted & Roher, 1994). That is, during a preset time, the individual recalls everything (s)he recalls about the material that was studied (words, texts), without following the order in which this information was presented.

In a study that used this approach, Roediger and Karpicke (2006a) investigated the testing effect and compared it with possible study effects. In this experiment, participants initially read short texts. Then, while one group freely recalled these texts, another group restudied them (reread the text an average three times). The final test (free recall) was applied after intervals of five minutes, two days and one week. On the immediate test (five minutes after the most recent reading), the restudy group performed better on the memory test than the recall group. After larger retention intervals (two and seven days) however, the results were inverted: the free recall group recalled more information from the original text than the restudy group. This result is quite surprising, considering that, as the level of exposure to the original text had actually been higher in the restudy group, the most intuitive prediction would be a better memory of the latter on the final recall test. Hence, this research convincingly demonstrated that an initial recall test, held immediately after the study, can improve the performance on final memory tests after long retention intervals (at least two days).

One important issue that results from these findings relates to the possibility of achieving this effect when educational material is used. To investigate this possibility, Karpicke and Roediger (2010) undertook two experiments in which they used encyclopedic texts as stimuli. The participants studied the texts and freely recalled them in an initial learning phase. To put the testing effect in practice, memory tests were conducted at varying intervals and the presence or absence of feedback was manipulated (supply of correct answer; in this case the renewed presentation of the excerpt). The findings showed that various repeated tests increased the retention more than a single test; tests that came with feedback produced a better retention than tests without feedback; and the intervals between the memory tests did not result in different testing effects. Hence, repeated retrieval through tests enhances long-term retention, but different test intervals do not serve as a determining factor.

The fact that a memory test series more strongly influences the long-term memory retention than repeated studies was initially demonstrated by Karpicke and Roediger (2007). These authors used three study and test arrangements in their experiments: standard (study-test-study-test), repeated study (study-study-study-test) and repeated test (study-test-test-test), using word lists as study material. One week after the participants had studied and been tested according to these

arrangements, they underwent a free recall test. The study found that repeated study and repeated test entailed different results on the final test: a series of tests is more beneficial than a series of studies for long-term retention purposes, supporting the assertion that retrieval is the key to promote memory and learning (Roediger & Karpicke, 2006a).

The occurrence of the testing effect in the free recall format was also demonstrated by Szpunar, McDermott and Roediger (2008). Their experiments checked whether tests can isolate the influence of proactive interferences (harmful interference for the learning of new information caused by previously memorized information) and suggested that test soon after having studied certain information protect against this type of interference that harms the establishment of long-term memories.

Besides protecting against interferences, recall can benefit memory retention because it improves the organization of information in specific categories. To investigate this possibility, Zaromb and Roediger (2010) used free recall measures and ranked word lists. As the authors expected, the memory tests produced further grouping in categories than the mere restudy of the stimuli and, the greater the organization in categories, the better the memory retrieval rates. These results evidence that organizational processes (that is, categorization processes) can play a fundamental role in the testing effect.

In general, based on the analysis of the seven papers presented in this section, it can be affirmed that free recall tests entail a clearly beneficial effect for long-term memory retention, as they convincingly demonstrate that recalling some information is more effective to retain it in the long-term memory than simply restudying that same information. This gives rise to the question: does this effect also take place when the recall is preceded by cues? Studies that attempted to answer this question will be discussed in the next section.

Cued recall tests. While the previous section discussed free recall experiments to verify the testing effect, in this section, the results of 13 studies are presented in which the testing effect is investigated using cued recall tests. These can be designed in different manners: through the learning of associated pairs (Carpenter, 2011; Carpenter, Pashler, & Vul, 2006; Carpenter, Pashler, Wixted, & Vul, 2008; Vojdanoska, Cranney, & Newell, 2010), short answers (Agarwal, Karpicke, Kang, Roediger, & McDermott, 2008; Butler, 2010) or by filling out gaps (McDaniel, Roediger, & McDermott, 2007).

When a person learns names for faces, the phone number of a friend, or that “coelho” means rabbit in Portuguese, the task fundamentally involves the learning of associated pairs (Allen et al., 1969; Estes, 1960; Jacoby, 1978). In general, in laboratory studies, the learning of associated pairs is studied by using word pairs that can vary in terms of their association value, or even by using pairs that include a “non-word” (“hitoyldace” for example) and a word.

The learning of associated word pairs was studied by Carpenter et al. (2006) for example, who investigated the testing effect through two experiments, using cued recall tasks. In experiment 1, the participants studied word pairs (A + B) and then took a cued recall test (A - ?) with feedback (the feedback involved showing the correct answer) or simply restudied the material (A + B). The final test took place after 24 hours and took the form of cued recalling (A - ? or ? - B) or free recalling of cues (A) or targets (B) only. All four final test formats revealed benefits of the test in comparison with the study, making the researchers conclude that cued retrieval (A - ?) followed by feedback (representation of the word pair: A + B) increases retention further than the same length of study time.

Retrieval through associated words was also investigated in the studies by Hays, Kornell and Bjork (2010), Karpicke and Bauernschmidt (2011) and Toppino and Cohen (2009). In the three studies, lists of foreign words were used. The participants learned word pairs in Swahili (one of the official language in Kenya, Tanzania and Uganda) and English through the repeated retrieval of the word in English when presented in combination with the word in Swahili. Hays et al. (2010) focused on the practice of feedback and concluded that it is not universally beneficial. Although usually positive, in some circumstances, providing feedback can negatively affect learning. The authors also demonstrated that, when the participants were authorized to control the feedback, they often ignored it when they considered it was necessary and displayed a better memory performance when they did so. When they received more time for the retrieval instead of feedback provision attempts, learning was also enhanced. The results emphasized the importance of considering the costs and benefits of learning improvement interventions and demonstrated that, although the feedback generally results in a better performance, this pattern cannot be generalized to all memory test situations.

Besides tests with associated word pairs, cued recall tests can also take the form of short answers. Butler (2010) outlined experiments in which the participants studied six short excerpts on different topics. Then, they repeatedly studied some of the excerpts or answered tests about them. One week later, the participants took a final test to assess the retention and transference of the information from the excerpts. The main findings support those of all other studies reported so far: repeated tests produce greater retention and transference in the final test in comparison with mere renewed studying. This study, however, broadened the results discussed earlier by demonstrating that the testing effect in cued recall tasks also takes place when texts are used as stimuli.

Cued recall tests in the form of short answers to text excerpts were also investigated in the experiments reported by Butler and Roediger (2007), Chan, McDermott and Roediger (2006) and Kornell, Hays and Bjork (2009). More specifically in the study by Agarwal et al. (2008), participants studied prose excerpts and then either restudied or took a

test with or without consulting the material studied. Tests with consulting improved the initial performance more than tests without consulting, but this benefit did not continue over time and both tests produced equivalent retention in tests taken after longer intervals. All studies referred to in this paragraph replicated the testing effect, using texts and excerpts as stimuli.

Another activity modality in the form of cued recalling is the completion of gaps. McDaniel et al. (2007) analyzed three experiments that extended the experimental results to educational material (short papers) and showed solid benefits of in-class testing through the completion of gaps. The results were similar to the word pair and short answer tests: initial cued recall tests significantly benefited the performance on the final test compared with not taking tests. Like the other forms, tests with the completion of gaps involved the production or recall of material and enhanced learning and long-term retention.

In short, the main conclusion of the cued recall studies, similar to the free recall studies, is that tests promote better retention when compared to study sessions. Next, we will verify whether this effect can also take place when recognition tests are used.

Studies About the Testing Effect Using Recognition Tests

In recognition tests, previously coded (memorized) items need to be distinguished from previously non-coded (new) items. Thus, in contrast when free and cued recall tests, in recognition tests, previously memorized information, like new information, is directly presented to the participants. The basic principle of these tests is to compare information presented during the test with information stored in the long-term memory, like in multiple-choice or true-or-false tests, which are frequently employed in educational contexts (Fazio, Agarwal, Marsh, & Roediger, 2010). Multiple-choice tests, addressed in the 11 studies discussed in this section, are frequently used in the classroom, considering that they are easier to score and their score is perceived as objective (Marsh, Agarwal, & Roediger, 2009; McDaniel, Agarwal, Huelser, McDermott, & Roediger, 2011; Roediger, Agarwal, McDaniel, & McDermott, 2011).

Although this kind of tests presents satisfactory long-term retention results, they may entail negative consequences for the memory: due to their exposure to incorrect alternatives, the students can store false facts in the long-term memory, mainly if they do not receive feedback after giving their answers (Butler, Marsh, Goode, & Roediger, 2006; Butler & Roediger, 2008; Marsh, Roediger, Bjork, & Bjork, 2007). Hence, the presence of different alternatives in multiple-choice tests can negatively affect the students' knowledge (Fazio et al., 2010; Roediger & Marsh, 2005). As demonstrated by Toppino and Luipersbeck (1993), when students judge the veracity of the information after a multiple-choice test, their performance increases for correct as well as incorrect answers. That is, when they consider a

wrong answer as correct in a multiple-choice test, individuals fundamentally recode this answer as if it were correct and replicate it in the final recall test.

Different studies have used recognition tests in the form of multiple-choice questions to verify the testing effect (Butler, Karpicke, & Roediger, 2007, 2008; Butler & Roediger, 2008; Carpenter, Pashler, & Cepeda, 2009; McDaniel, Howard, & Einstein, 2012). The studies cited in this review demonstrated that the beneficial effects of the tests continue after a certain period, but did not investigate the durability of the tests' negative (or harmful) effects, such as the recalling of incorrect items. Thus, Fazio et al. (2010) conducted an experiment in which the participants solved multiple-choice tests and cued recall tests, both administered immediately and one week after the study session (non-fiction text excerpts). The results demonstrated that the interval reduces both the positive and negative testing effects.

Also aiming to investigate the testing effect based on multiple-choice tests, Butler et al. (2006) conducted three experiments to investigate whether increasing the number of new items in multiple-choice tests further harms the memory performance. The results indicated that the performance on the initial multiple-choice test was an important factor: when the performance on this test was very high, the increase in the number of new items entailed a better performance on subsequent tests; on the other hand, in case of lower performance levels on the initial test, new additional items negatively affected further test scores.

One possible solution to minimize the negative effects of multiple-choice tests can be to offer feedback. Thus, Butler and Roediger (2008) investigated whether feedback can be used to enhance the positive effects and reduce the negative effects of multiple-choice tests. Therefore, the participants studied excerpts from historical topics and took an initial multiple-choice test with immediate feedback, later feedback and no feedback. The results suggested that both immediate feedback and later feedback increased the proportion of correct answers and reduced the proportion of intrusions, regarding incorrect alternatives presented in the initial multiple-choice test.

Based on a review of experiments using educational materials, such as short papers and college course syllabi, McDaniel et al. (2007) showed that short-answer tests are more effective than multiple-choice tests. It is important, however, to know the effects of using multiple-choice tests on successful information retrieval, considering that these are frequently used in the classroom. In short, the reviewed studies that applied recognition tests support the findings of studies that used recall tests. Evidence, however, points towards the conclusion that tests involving participants' production of information (free or cued recall tests) produce greater benefits than multiple-choice tests, which merely involve the recognition of the correct answer among different options.

Discussion

In the previous sections, recent experiments were summarized about the test formats used to investigate the testing effect and robust findings were described about the memory performance, which can be extended to educational contexts and materials. The results specifically addressed a range of designs and contexts, like basic laboratory work with word lists, laboratory experiments with educative materials and different retention intervals and in-class experiments. The results indicated that: (1) initial short-answer or multiple-choice tests significantly benefitted participants' performance on subsequent tests (in comparison with not taking tests); (2) initial short-answer tests (production or recall of material) benefit performance on subsequent tests more than multiple-choice (recognition) tests, even when the final tests adopt a multiple-choice design, and (3) the benefits of short-answer, but not necessarily of multiple-choice tests, significantly exceed the study of the target material (McDaniel et al., 2007). Thus, free recall tests entail more positive effects for long-term retention than recognition tests, but both are more beneficial when compared to simple study (rereading of the material). These effects are even more prominent when the initial tests include corrective feedback.

Hence, due to the noteworthy memory effect of the tests discussed earlier, investigating the testing effect in the classroom environment arouses increasing interest. The interest in the application of experimental research on memory in the educational context is aimed at enhancing knowledge acquisition and retention in the school environment (Carpenter et al., 2009; McDaniel et al., 2011). The current stage of research on the testing effect in the classroom reflects the translation of findings from the laboratory to the classroom environment and already shows concrete applications of this effect in the educational sphere. In an analysis of three experiments that extended their results to relevant educational material, such as text excerpts, McDaniel et al. (2007) demonstrated robust benefits resulting from the testing effect in the classroom as well.

This review also revealed that the participants' age range has been hardly explored. Most studies on the testing effect involved adult college students. Recently, some studies have been concerned with involving younger age ranges, like Carpenter et al. (2009), in which the retention of US historical facts by eighth-grade students was assessed. Rohrer et al. (2010) investigated the testing effect in nine and ten-year-old children to verify differences between final test types: one with questions exactly equal to the initial test questions, and another with more challenging questions that demanded transference. Bouwmeester and Verkoeijen (2011) studied seven to 13-year-olds, investigating whether individual differences in the testing effect can be attributed to variations in the processing of the memory trait, using semantically associated word lists (DRM lists). McDaniel et al. (2011) investigated the testing effect in eighth-grade students, involving multiple-choice tests. Even in experiments that involved participants from different age

ranges, the testing effect was verified. Research on the testing effect in the classroom is a promising research approach, which can offer countless practical benefits to Education, as well as for the understanding of basic memory processes in ecologically valid situations.

One of the objectives of Education, perhaps the most fundamental one, is to maximize long-term knowledge retention and transfer (Hays et al., 2010). As demonstrated in the papers discussed in this review, a key approach to achieve this objective is to promote the retrieval of information for learning through periodical tests. One method to enhance the benefits of tests and reduce the negative effects of students' exposure to wrong information (like when memorizing an incorrect alternative in a multiple-choice test) is to provide feedback (Butler & Roediger, 2008). Recent studies on the effects of feedback time have suggested that this technique can also reinforce correct answers (Smith & Kimball, 2010). In the papers analyzed, it was observed that the testing effect is robust, even without feedback, but that, when it is offered to the participants, this effect tends to be more significant.

Experimental findings about the testing effect are also important to understand basic memory processes. In general, however, the papers analyzed described or superficially speculated on what the testing effect represents in cognitive terms. The results showed that successfully retrieving a piece of information is more beneficial than simply restudying the target material, but further theorization is lacking on what this effect represents in terms of cognitive processes. Some questions still need further explanation: does the testing effect result from easier access to the information from the original material? Or is the participants' better performance due to the fact that the tasks (taking one more test) are familiar?

In general, the theoretical explanations for the effects of retrieval (tests) were concentrated on how the act of reminding affects the memory performance. Retrieving a piece of information from memory leads to the elaboration of the memory trait and/or the creation of additional retrieval routes (Carpenter, 2009; Roediger & Butler, 2011), making it more probable that the memory will be retrieved successfully in the future. McDaniel and Fisher (1991) and McDaniel and Masson (1985) affirmed that the tests improve learning because they generate the elaboration of existing memory traits and the elaboration of the cue-target relations, that is, the relation between the memory cue and the memory trait associated with that cue. Thus, according to that interpretation, the tests multiply the number of retrieval routes for stored events (Roediger & Karpicke, 2006b).

Two other hypotheses related to the cognitive processing that underlies the testing effect are frequently considered. The first refers to the effort made when memories are retrieved, that is, the amount of reprocessing of the memory trait that takes place during the retrieval (Pyc & Rawson, 2009). According to this hypothesis, the more effort is involved in retrieving a memory, the more extensive the reprocessing of the memorized information

will be. Thus, tests that make it more difficult to recall are more beneficial to long-term retention (McDaniel et al., 2007). The second hypothesis is based on the transfer-appropriate processing concept, which defends that memory performance is reinforced to the extent that the cognitive processes required during learning coincide with those required during retrieval (Roediger, 1990). Transfer-appropriate processing emphasizes compatibility between the operations involving during the test and learning phase; the processes involved in taking an initial test are more appropriate for the final test than the processes involved in studying the material (Roediger & Butler, 2011).

The testing effect is a powerful means to improve learning and long-term retention. Its importance is renowned in the educational area and in teaching and learning Psychology. The practical implication of the present review results is that students should be tested (or self-test) repeatedly while learning, not only because tests provide knowledge of the results which can direct future study, but also because information retrieval strongly benefits retention. On the other hand, the studies reviewed come with limitations. The testing effect is mainly assessed through memory tests (recall and recognition), with faint demonstrations of how this effect can be extrapolated to the different assessment forms teachers use in the teaching-learning process. Even if repeated retrieval is a powerful manner of promoting retention, it is not clear that students and educators consider taking a test as a learning device or use it as a tool to enhance learning. Future studies in the area can provide information on the types of study techniques recommended for students and educators. It is equally important to discuss the influence of the testing effect beyond the classroom environment, when the same assessment instruments are used to measure the efficacy of a certain treatment type, to develop psychological assessments for the verification of performance tests, and to undertake health research and practice with repeated applications of cognitive neuropsychological tests to assess the effectiveness of psychosocial or pharmacological treatments.

Final Considerations

In this systematic literature review, it was verified that the testing effect (that is, increased and improved long-term retention and learning through tests) manifests itself in the form of different patterns. This is a recurrent effect in different populations (children, adolescents, young adults), which involves various materials (lists of words or images, texts) and retention intervals that range from hours to days, a week or a month. The testing effect is extremely robust since, as demonstrated in this review, it has been replicated in dozens of studies.

The results of research about this effect entail direct implications for the optimization of learning methods and instructions used in the classroom. Although highly popular

in the educational context, the mere restudying of certain contents seems to be a hardly effective learning technique. On the opposite, teaching methods that encourage the retrieval of previously memorized information can be very useful for successful retention in the long term and, consequently, for learning. Therefore, in educational contexts, it is important to use the advantages this effect offers to benefit learning.

References

- Agarwal, P. K., Karpicke, J. D., Kang, S. H. K., Roediger, H. L., III, & McDermott, K. B. (2008). Examining the testing effect with open- and closed-books tests. *Applied Cognitive Psychology*, 22(7), 861-876. doi:10.1002/acp.1391
- Allen, G. A., Mahler, W. A., & Estes, W. K. (1969). Effects of recall tests on long-term retention of paired associates. *Journal of Verbal Learning and Verbal Behavior*, 8(4), 463-470. doi:10.1016/S0022-5371(69)80090-3
- Anderson, J. R. (2005). *Aprendizagem e memória: Uma abordagem integrada* (J. A. Saad, Trad., 2nd ed.). Rio de Janeiro: LTC.
- Blumen, H. M., & Stern, Y. (2011). Short-term and long-term collaboration benefits on individual recall in younger and older adults. *Memory & Cognition*, 39(1), 147-154. doi:10.3758/s13421-010-0023-6
- Bouwmeester, S., & Verkoeijen, P. P. J. L. (2011). Why do some children benefit more from testing than others? Gist trace processing to explain the testing effect. *Journal of Memory and Language*, 65(1), 32-41. doi:10.1016/j.jml.2011.02.005
- Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(5), 1118-1133. doi:10.1037/a0019902
- Butler, A. C., Karpicke, J. D., & Roediger, H. L., III. (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology: Applied*, 13(4), 273-281. doi:10.1037/1076-898X.13.4.273
- Butler, A. C., Karpicke, J. D., & Roediger, H. L., III. (2008). Correcting a metacognitive error: Feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(4), 918-928. doi:10.1037/0278-7393.34.4.918
- Butler, A. C., Marsh, E. J., Goode, M. K., & Roediger, H. L., III. (2006). When additional multiple-choice lures aid versus hinder later memory. *Applied Cognitive Psychology*, 20(7), 941-956. doi:10.1002/acp.1239
- Butler, A. C., & Roediger, H. L., III. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology*, 19(4-5), 514-527. doi:10.1080/09541440701326097

- Butler, A. C., & Roediger, H. L., III. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition*, 36(3), 604-616. doi:10.3758/MC.36.3.604
- Callender, A. A., & McDaniel, M. A. (2009). The limited benefits of rereading educational texts. *Contemporary Educational Psychology*, 34(1), 40-41. doi:10.1016/j.cedpsych.2008.07.001
- Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(6), 1563-1569. doi:10.1037/a0017021
- Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(6), 1547-1552. doi:10.1037/a0024140
- Carpenter, S. K., & Pashler, H. (2007). Testing beyond words: Using tests to enhance visuospatial map learning. *Psychonomic Bulletin & Review*, 14(3), 474-478. doi:10.3758/BF03194092
- Carpenter, S. K., Pashler, H., & Cepeda, N. J. (2009). Using tests to enhance 8th grade students' retention of U.S. history facts. *Applied Cognitive Psychology*, 23(6), 760-771. doi:10.1002/acp.1507
- Carpenter, S. K., Pashler, H., & Vul, E. (2006). What types of learning are enhanced by a cued recall test? *Psychonomic Bulletin & Review*, 13(5), 826-830. doi:10.3758/BF03194004
- Carpenter, S. K., Pashler, H., Wixted, J. T., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition*, 36(2), 438-448. doi:10.3758/MC.36.2.438
- Chan, J. C. K., McDermott, K. B., & Roediger, H. L., III. (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General*, 135(4), 553-571. doi:10.1037/0096-3445.135.4.553
- Congleton, A. R., & Rajaram, S. (2011). The influence of learning methods on collaboration: Prior repeated retrieval enhances retrieval organization, abolishes collaborative inhibition, and promotes post-collaborative memory. *Journal of Experimental Psychology: General*, 140(4), 535-551. doi:10.1037/a0024308
- Estes, W. K. (1960). Learning theory and the new "mental chemistry". *Psychological Review*, 67(4), 207-223. doi:10.1037/h0041624
- Fazio, L. K., Agarwal, P. K., Marsh, E. J., & Roediger, H. L., III. (2010). Memorial consequences of multiple-choice testing on immediate and delayed tests. *Memory & Cognition*, 38(4), 407-418. doi:10.3758/MC.38.4.407
- Halamish, V., & Bjork, R. A. (2011). When does testing enhance retention? A distribution-based interpretation of retrieval as a memory modifier. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(4), 801-812. doi:10.1037/a0023219
- Hays, M. J., Kornell, N., & Bjork, R. A. (2010). The costs and benefits of providing feedback during learning. *Psychonomic Bulletin & Review*, 17(6), 797-801. doi:10.3758/PBR.17.6.797
- Hogan, R. M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning and Verbal Behavior*, 10(5), 562-567. doi:10.1016/S0022-5371(71)80029-4
- Howe, M. L. (1991). Misleading children's story recall: Forgetting and reminiscence of the facts. *Developmental Psychology*, 27(5), 746-762. doi:10.1037/0012-1649.27.5.746
- Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior*, 17(6), 649-667. doi:10.1016/S0022-5371(78)90393-6
- Jacoby, L. L., Wahlheim, C. N., & Coane, J. H. (2010). Test-enhanced learning of natural concepts: Effects on recognition memory, classification, and metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(6), 1441-1451. doi:10.1037/a0020636
- Karpicke, J. D. (2012). Retrieval-based learning: Active retrieval promotes meaningful learning. *Current Directions in Psychological Science*, 21(3), 157-163. doi:10.1177/0963721412443552
- Karpicke, J. D., & Bauernschmidt, A. (2011). Spaced retrieval: Absolute spacing enhances learning regardless of relative spacing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(5), 1250-1257. doi:10.1037/a0023436
- Karpicke, J. D., & Roediger, H. L., III. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, 57(2), 151-162. doi:10.1016/j.jml.2006.09.004
- Karpicke, J. D., & Roediger, H. L., III. (2010). Is expanding retrieval a superior method for learning text materials? *Memory & Cognition*, 38(1), 116-124. doi:10.3758/MC.38.1.116
- Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(4), 989-998. doi:10.1037/a0015729
- Marsh, E. J., Agarwal, P. K., & Roediger, H. L., III. (2009). Memorial consequences of answering SAT II questions. *Journal of Experimental Psychology: Applied*, 15(1), 1-11. doi:10.1037/a0014721
- Marsh, E. J., Roediger, H. L., III, Bjork, R. A., & Bjork, E. L. (2007). The memorial consequences of multiple-choice testing. *Psychonomic Bulletin and Review*, 14(2), 194-199. doi:10.3758/BF03194051

- McDaniel, M. A., Agarwal, P. K., Huelser, B. J., McDermott, K. B., & Roediger, H. L., III. (2011). Test-enhanced learning in a middle school science classroom: The effects of quiz frequency and placement. *Journal of Educational Psychology*, 103(2), 399-414. doi:10.1037/a0021782
- McDaniel, M. A., & Fisher, R. P. (1991). Tests and test feedback as learning sources. *Contemporary Educational Psychology*, 16(2), 192-201. doi:10.1016/0361-476X(91)90037-L
- McDaniel, M. A., Howard, D. C., & Einstein, G. O. (2012). The read-recite-review strategy: Effective and portable. *Psychological Science*, 20(4), 516-522. doi:10.1111/j.1467-9280.2009.02325.x
- McDaniel, M. A., & Masson, M. E. (1985). Altering memory representation through retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11(2), 371-385. doi:10.1037/0278-7393.11.2.371
- McDaniel, M. A., Roediger, H. L., III, & McDermott, K. B. (2007). Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin & Review*, 14(2), 200-206. doi:10.3758/BF03194052
- Moreira-Aguiar, V., Lameira, A. P., Conde, E. Q., Pereira Júnior, A., Umiltà, C. A., & Gawryszewski, L. G. (2008). Memória de longo prazo modulada pela memória de curto prazo. *Paidéia (Ribeirão Preto)*, 18(40), 331-339. doi:10.1590/S0103-863X2008000200010
- Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory & Cognition*, 31(1), 3-8. doi:10.1037/0278-7393.31.1.3
- Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language*, 60(4), 437-447. doi:10.1016/j.jml.2009.01.004
- Roediger, H. L., III. (1990). Implicit memory: Retention without remembering. *American Psychologist*, 45(9), 1043-1056. doi:10.1037/0003-066X.45.9.1043
- Roediger, H. L., III, Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (2011). Test-enhanced learning in the classroom: Long-term improvements from quizzing. *Journal of Experimental Psychology: Applied*, 17(4), 382-395. doi:10.1037/a0026252
- Roediger, H. L., III, & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, 15(1), 20-27. doi:10.1016/j.tics.2010.09.003
- Roediger, H. L., III, & Karpicke, J. D. (2006a). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17(3), 249-255. doi:10.1111/j.1467-9280.2006.01693.x
- Roediger, H. L., III, & Karpicke, J. D. (2006b). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, 1(3), 181-210. doi:10.1111/j.1745-6916.2006.00012.x
- Roediger, H. L., III, & Marsh, E. J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(5), 1155-1159. doi:10.1037/0278-7393.31.5.1155
- Roediger, H. L., III, McDermott, K. B., & McDaniel, M. A. (2011). Using testing to improve learning and memory. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 65-74). New York: Worth.
- Rohrer, D., Taylor, K., & Sholar, B. (2010). Tests enhance the transfer of learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 36(1), 233-239. doi:10.1037/a0017678
- Smith, T. A., & Kimball, D. R. (2010). Learning from feedback: Spacing and the delay-retention effect. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 36(1), 80-95. doi:10.1037/a0017407
- Szpunar, K. K., McDermott, K. B., & Roediger, H. L., III (2008). Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(6), 1392-1399. doi:10.1037/a0013082
- Toppino, T. C., & Cohen, M. S. (2009). The testing effect and the retention interval: Questions and answers. *Experimental Psychology*, 56(4), 252-257. doi:10.1027/1618-3169.56.4.252
- Toppino, T. C., & Luipersbeck, S. M. (1993). Generality of the negative suggestion effect in objective tests. *Journal of Educational Psychology*, 86(6), 357-362. doi:10.1080/00220671.1993.9941229
- Vojdanoska, M., Cranney, J., & Newell, B. R. (2010). The testing effect: The role of feedback and collaboration in a tertiary classroom setting. *Applied Cognitive Psychology*, 24(8), 1183-1195. doi:10.1002/acp.1630
- Watkins, M. J. (1974). The concept and measurement of primary memory. *Psychological Bulletin*, 81(10), 695-711. doi:10.1037/h0036952
- Weinstein, Y., McDermott, K. B., & Szpunar, K. K. (2011). Testing protects against proactive interference in face-name learning. *Psychonomic Bulletin & Review*, 18(3), 518-523. doi:10.3758/s13423-011-0085-x
- Wixted, J. T., & Rohrer, D. (1994). Analyzing the dynamics of free recall: An integrative review of the empirical literature. *Psychonomic Bulletin & Review*, 1(1), 89-106. doi:10.3758/BF03200763

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, 46(3), 441-517. doi:10.1006/jmla.2002.2864

Zaromb, F. M., & Roediger, H. L., III (2010). The testing effect in free recall is associated with enhanced organizational processes. *Memory & Cognition*, 38(8), 995-1008. doi:10.3758/MC.38.8.995

Raquel Eloísa Eisenkraemer is a Ph.D. in Psychology from Pontifícia Universidade Católica do Rio Grande do Sul.

Antônio Jaeger is a Postdoctoral Researcher of the Universidade de São Paulo.

Lilian Milnitsky Stein is a Full Professor of the Pontifícia Universidade Católica do Rio Grande do Sul.

Received: Jul. 26th 2012

1st Revision: Feb. 15th 2013

2nd Revision: Apr. 24th 2013

Approved: May 7th 2013

How to cite this article:

Eisenkraemer, R. E., Jaeger, A., & Stein, L. M. (2013). A systematic review of the testing effect in learning. *Paidéia (Ribeirão Preto)*, 23(56), 397-406. doi:10.1590/1982-43272356201314